12-2007

Creating alternate in-basket forms through cloning: Some preliminary results

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Frederik ANSEEL
Ghent University

DOI: https://doi.org/10.1111/j.1468-2389.2007.00401.x

Creating Alternate In-Basket Forms Through Cloning: Some preliminary results

Filip Lievens and Frederik Anseel

Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri, Dunantlaan 2, 9000 Ghent, Belgium. filip.lievens@ugent.be

Published in International Journal of Selection and Assessment, 2007 December, 15 (4), 428-433.

https://doi.org/10.1111/j.1468-2389.2007.00401.x

Submitted version

Abstract

Research on constructing alternate forms of assessment center exercises is very scarce. This study examines the effectiveness of a cloning procedure (incident isomorphic approach) for developing alternate forms of a computerized in-basket. In this approach, original and alternate items are essentially similar (they are based on the same critical incident), while being superficially different (they are situated in a different context). Results showed there was no significant difference between the overall in-basket score across the alternate forms. In addition, these overall scores correlated .66, with projected estimates for the full in-basket approaching .80. Implications and limitations of the use of cloning in designing alternate assessment center exercises are discussed.

There are various reasons as to why developing alternate forms of assessment center exercises is useful and important. First, the use of alternate forms is important from a legal point of view. If candidates are given exercises that are not similar, they might take legal action, appealing that there was no equal treatment across different test administrations. Second, the use of alternate forms might reduce candidates' overexposure to specific assessment exercises. In fact, repeated use of a given assessment center exercise within a given selection process or across various selection processes might result in this exercise being known by applicants. Third, the design of alternate forms of a given assessment center exercise seems less costly than constructing an entirely new exercise.

This study contributes to the scarce literature on exercise design in assessment centers (Ahmed, Payne, & Whiddett, 1997; Thornton & Mueller-Hanson, 2004) by presenting and examining a possible approach for developing alternate forms of assessment center exercises. Specifically, we borrow from cloning procedures that were recently developed for situational judgment tests (Clause, Mullins, Nee, Pulakos, & Schmitt, 1998; Lievens & Sackett, 2007) and apply them to one of the most widely used assessment center exercises (i.e., a computerized in-basket exercise). In cloning procedures, alternate test items are constructed that look superficially differently, even though they are psychometrically equivalent. In other words, the objective of cloning consists in changing the surface characteristics of items that do not

determine item difficulty (also referred to as incidentals), while holding constant the structural item features that determine item difficulty (also referred to as radicals, see Irvine & Kyllonen, 2002).

2. Constructing alternate forms of assessment center exercises

Domain sampling has been the traditional approach for constructing alternate forms of tests (Nunnally & Bernstein, 1994). In the domain-sampling approach, items are sampled from a larger existing bank of items whose underlying constructs and psychometric properties [e.g., item mean, item standard deviation (SD)] are known. It is well documented that this domain-sampling model is effective for creating alternate forms of tests, which purport to measure constructs that are well understood. Examples are achievement tests, knowledge tests, general cognitive ability tests, and specific cognitive ability tests (see Armstrong, Jones, & Wang, 1994; Gibson & Weiner, 1998).

The domain-sampling approach has also been used in assessment centers, especially for developing alternate in-basket forms. However, results have been disappointing. For example, Schippmann, Prien, and Katz (1990) reviewed various earlier studies that examined the alternate-form reliability of in-basket exercises. They reported low alternate-form reliability coefficients varying from.15 to.38. In another study, Brannick, Michaels, and Baker (1989) examined the reliability of two in-baskets developed for the same position in a research context. Alternate-form reliabilities ranged from.21 to.43. Brannick et al. also randomly split the two in-baskets into two halves (odd and even items). Even in that case, alternate-form reliabilities were low.

These disappointing results for the domain-sampling approach are not surprising. Clause et al. (1998) cogently argued that the domain-sampling approach will not work for multidimensional tests or tests whose constructs are poorly understood. The problem seems to be that even relatively minor changes in the situations presented in these multidimensional predictors might result in other constructs being elicited, resulting in the lack of consistency across forms.

In light of these limitations of the domain-sampling approach for constructing alternate forms of multidimensional tests, a cloning procedure has been suggested as a possible alternative (Clause et al., 1998; Lievens & Sackett, 2007). Generally, two cloning approaches can be distinguished. An item isomorphic approach maximizes the degree of similarity between the original and alternate test forms by constructing similar items in terms of critical incident embedded, context of the incident, and grammatical structure as the items of the original test. Essentially, this means that there are only cosmetic differences between the items of the respective forms. Clause et al. used this approach to construct alternate forms of a situational judgment test. They examined the viability of this item isomorphic approach procedure in a research context and found alternate-form reliabilities above.70 for a 33-item situational judgment test. An incident isomorphic approach is less restrictive as the degree of similarity between the original and alternate test forms is ensured by constructing similar items in terms of the critical incident on which the item is based (Lievens & Sackett, 2007). However, the items differ in terms of the specific context wherein the critical incident is embedded.

We believe that these cloning approaches might also be useful for constructing alternate forms of the problem descriptions used in assessment center exercises. This is because assessment center exercises are also measurement methods that can be designed to capture various constructs (Arthur, Day, McNelly, & Edens, 2003; Lievens, Harris, Van Keer, & Bisqueret, 2003). In addition, the constructs measured by assessment centers are often poorly understood and seem to be contingent upon the situations sampled (Woehr & Arthur, 2003). Therefore, we applied a cloning procedure (incident isomorphic) to develop alternate forms of a computerized in-basket. We chose the incident isomorphic approach because it is less restrictive than the item isomorphic approach. The incident isomorphic approach ensures the similarity across forms (as indicated by the correlation among the two test forms), while keeping practice effects (as indicated by mean score changes due to prior exposure to the original test form) within bounds because original and alternate items are essentially similar (they are based on the same critical incident), while being superficially different (they are situated in a different context).

3. Method

3.1. Sample and procedure

We examined the effectiveness of our cloning procedure in a research and field setting. Hence, two samples were used. The first sample consisted of 295 final-year students (68% females, 32% males) from different majors (e.g., engineering, agricultural sciences, communication, education). Participants had an average age of 22.2 years (SD=1.6 years). The student sample participated in exchange for course credit. Sessions were conducted in groups of 45 individuals in large PC-equipped rooms. To increase involvement, participants were told that in-baskets are often used in selection procedures for junior managers. Participants were also informed that they would receive feedback about their performance at the end.

The second sample consisted of 485 participants1 (56% male, 44% female). Their ages ranged from 18 to 59 years (M=36.10 years, SD=9.7 years). They had an average work experience of 13.3 years (SD=9.9 years). The four largest job categories were clerical jobs (17.4%), technical work (16.6%), service jobs (15.3%), and executive positions (11.2%). The remaining job categories included production, education, trades, and sales. About 9.7% of the sample reported to be currently unemployed. Seventy-five percent held at least a bachelor's degree and 40% had earned an advanced or professional degree. This sample was recruited by placing a link to the in-basket on the website of the Flemish governmental service for employment and vocational training (VDAB). This website is frequently visited by applicants and employees who look for training and coaching in job application skills and work-related competencies. The in-basket was advertised as a way to prepare oneself for potential future selection procedures and for assessing one's managerial competencies. Upon completion of a short questionnaire measuring demographic variables, people received a random respondent identifier that gave access to the website with the in-basket.

In both samples, participants completed both the original and the cloned in-basket forms. To avoid order and practice effects, these two forms were presented in a counterbalanced order. Participants took about 1

h to complete both in-basket forms. Afterwards, they received feedback on their overall in-basket score and score on each of the four competencies measured.

3.2. Development of alternate in-basket forms

We used an adapted version of an in-basket exercise that was originally developed by Tett and Jackson (1990). The original in-basket comprised of 20 memos and letters addressed to a manager of a fictitious paints manufacturing plant. The computerized version of this in-basket (Anseel & Lievens, 2006; Tett, Menard, Guterman, & Beauregard, 2001) simulated as closely as possible the key features of an actual email software program (e.g., opportunity to read and respond to emails in any order, ongoing access to organizational charts and background information, etc.). Upon opening an email message, participants were provided with four e-mail responses (i.e., four response options). Next, they indicated the effectiveness of each response option, ranging from 1=very ineffective to 5=very effective. Each response option measured one of four competencies included in a taxonomy of managerial competence, namely Problem Awareness, Coordinating, Information Management, and Decisiveness (Tett, Guterman, Bleier, & Murphy, 2000).

To construct an alternate form of the computerized in-basket, we decided to clone only 10 of the 20 items because otherwise the in-basket would be too time consuming for the participants. For each original item, we wrote a new item including the same problem situation (critical incident) as the original item. However, on the surface the item looked different as it was situated in a different context. Three independent experts rated the degree of content similarity between the original and cloned items (item stems and item options) on a seven-point scale, ranging from not at all similar in terms of content to very similar in terms of content (κ=.84). The average similarity rating across all items was 6.36 (SD=.71). The end result was that the in-basket contained ten original and 10 cloned items. By way of an example, an original item is presented in Appendix A and a cloned item in Appendix B.

4. Results

Table 1 presents the means, SDs, and correlations among the two in-basket forms in the student sample. Effect sizes are also presented. These d values were obtained by subtracting the score on the second form by the score on the first form, divided by the pooled SD (Hunter & Schmidt, 2004, p. 277, equation 7.4). Positive effect sizes mean that the second form score was higher than the first one and are indicative of practice effects. We conducted a MANOVA with the overall in-basket score as dependent variable. In-basket form served as a within-subject factor and order as a between-subject factor. There was no significant main effect of in-basket form, $F(1, 292)=.36$, Wilks' $\lambda=.999$. The main effect of order and the interaction effect of in-basket form and order were also not significant. Although Table 1 shows some practice effects across forms with regard to the dimensional in-basket scores (d's ranging from −.30 to.36), they cancel each other out with respect to the overall score (d=.03).

Table 1. Means, standard deviations (*SD*s), and intercorrelations among alternate in-basket forms in student sample ($N = 294$)

	M	*SD*	*d*	1.	2.	3.	4.	5.	6.	7.	8.	9.
First form												
1. Coordinating	26.90	2.89	–	–								
2. Decisiveness	34.71	3.66	–	.30	–							
3. Information Management	37.88	3.33	–	.07	−.02	–						
4. Problem Awareness	47.12	4.35	–	.14	.20	.28	–					
5. Overall score	146.61	8.78	–	.55	.61	.53	.73	–				
Second form												
6. Coordinating	27.92	2.81	.36	.36	.21	.04	.16	.30	–			
7. Decisiveness	33.57	4.05	−.30	.28	.46	−.02	.26	.41	.30	–		
8. Information Management	38.31	3.52	.13	.12	.13	.46	.34	.44	.07	.19	–	
9. Problem Awareness	47.11	4.42	.00	.22	.16	.25	.65	.55	.17	.29	.43	–
10. Overall score	146.91	9.86	.03	.36	.37	.28	.57	.66	.51	.69	.64	.76

The correlations of the dimensional in-basket scores varied from.36 (Coordinating) to.65 (Problem Awareness). Logically, the alternate-form reliability coefficient (.66) was highest for the overall in-basket score. Note that these alternate-form reliability coefficients are based only on 10 in-basket items. Application of the Spearman–Brown prophecy formula projects the alternate-form reliability of the overall in-basket score to.79 for the full (20-item) in-basket.

Table 2 shows the means, SDs, and correlations among the two in-basket forms in the field sample. Of the total group of 485 individuals participating via the WWW, 69 stopped after completing the first form, reducing the sample to 416 participants. Therefore, we corrected the alternate-form reliability coefficients in Table 2 for direct-range restriction.

Table 2. Means, standard deviations (*SD*s), and intercorrelations among alternate in-basket forms in field sample ($N = 407$)

	M	*SD*	*d*	1	2.	3.	4.	5.	6.	7.	8.	9.
First form												
1. Coordinating	28.08	3.03	–	–								
2. Decisiveness	34.99	4.16	–	.27	–							
3. Information Management	37.56	3.54	–	.19	.15	–						
4. Problem Awareness	47.19	4.18	–	.13	.23	.30	–					
5. Overall score	147.81	9.59	–	.56	.68	.63	.68	–				
Second form												
6. Coordinating	28.24	3.03	.05	.36	.21	.20	.11	.33	–			
7. Decisiveness	34.98	4.41	.00	.36	.45	.19	.29	.51	.34	–		
8. Information Management	38.34	3.58	.22	.22	.21	.47	.27	.46	.26	.32	–	
9. Problem Awareness	46.68	4.67	−.12	.24	.23	.31	.57	.54	.26	.44	.32	–
10. Overall score	148.24	11.13	.04	.41	.40	.41	.47	.66	.60	.77	.65	.77

Generally, results of the field sample confirmed those of the student sample. Again, the MANOVA did not yield significant effects for in-basket form, order, and their interaction effect. As shown in Table 2, the difference between the overall in-basket scores across the forms was again minor (d=.04), indicating that there was no practice effect. There were small differences across forms with regard to the dimensional in-basket scores (d's ranging from −.12 to.22). The alternate-form reliability coefficient equaled.66 for the overall in-basket score. Using the Spearman–Brown prophecy formula, the alternate-form reliability of the overall in-basket score is projected to equal.80 for the full (20-item) in-basket.

5. Discussion

This study provided some preliminary evidence that cloning procedures (in this case an incident isomorphic approach) can be used for constructing alternate forms of a computerized in-basket. In both the student and field sample, there was no significant difference between the overall in-basket score across the alternate forms. In addition, these overall scores correlated.66, with projected estimates for the full in-basket approaching.80. All of this extends the positive results found for the use of cloning approaches for creating alternate forms of situational judgment tests (Clause et al., 1998; Lievens & Sackett, 2007). As we examined the similarity across the in-basket forms in a research and operational vocational guidance context, future research is needed to test this approach in actual (high stakes) selection.

In this study, we applied a cloning approach to both the item stems (problem situations) and response options of a computerized in-basket. However, we believe that such a cloning procedure might also be fruitfully used for developing alternate versions of the problem descriptions in paper-and-pencil in-baskets. In addition, future studies should examine the usefulness of cloning for constructing alternate forms of the problem descriptions of other assessment center exercises (role-plays, case-analyses, presentations, and group discussions) and in situational interviews. To this end, future research might also compare the effectiveness of item isomorphic and incident isomorphic approaches (Lievens & Sackett, 2007).

Although the development of alternate forms of assessment center exercises might increase the 'shelf life' of these exercises to some extent, it is clear that cloning assessment center exercises is at best a temporary solution. Original and cloned items are based on the same incident. Hence, using cloned items (especially if an item isomorphic approach is used) for a longer time period will lead to overexposure of the items. When test preparation firms know that cloning is applied, it seems also easy to prepare applicants for it. Therefore, future research should further develop and refine our existing techniques of exercise design in assessment centers.

Note

As this in-basket was administered through the WWW, we followed recommendations of Stanton and Rogelberg (2001) and screened the data (i.e., for responses not matching ‘legal’ identifiers and for inadvertent and malicious multiple responses). All of the suspect cases (about 10%) were removed.

Appendix A

Example of original computerized in-basket item

Date: Tuesday, 06/17/02 From: Eric Danforth

Subject: Upcoming Conference

Patrick,

Last week Mr Green arranged a meeting for July 2 regarding a new line of wind-resistant paints for use in the aviation industry. I realize my input would be valuable at this stage, but I've been planning since last fall to attend a conference in Boise, Idaho July 1–3 on new paint manufacturing processes. Is there any way the meeting could be postponed until after the 7th?

A. It's rather difficult to postpone this meeting as Mr Green is coming to our plant to attend this meeting. Is it possible to cancel your trip to Boise? Let's take some time to consider, I'll ask Mr Green and let you know what he thinks.

B. Indeed, it is probably more important for our company that you will attend the conference in Boise. If I understand correctly, the conference deals with the same innovativeness topic as our meeting. So, we will benefit from hearing the results of the conference and postpone our meeting.

C. I do not think the Boise-conference is that important. It's mostly advertising for larger companies. I wouldn't worry too much about attending the conference. I think it's better to proceed with our meeting as planned.

D. I can see that you have been planning this conference for some time now and that it takes priority. I will postpone the meeting to July 14th. Mr Green is aware of our priorities and won't make a problem of this.

Appendix B

Example of cloned computerized in-basket item

Date: Thursday, 06/19/02 From: Glen Benning

Subject: Exhibition in Chicago

Patrick,

You'll probably know we have arranged a meeting with all production supervisors for July 15th. We agreed to discuss in this meeting how we might improve customized production of industrial coatings. However, at the same day, there's an exhibition in Chicago on coating applications in the building industry, which I like to attend. I wondered if it would be possible to schedule our meeting at another day?

A. It's kind of hard to reschedule this meeting as all production supervisors have already confirmed their presence. Have you already made appointments with sales people at that exhibition? I'll first check with all other supervisors and then get back to you. So, we'll have to wait and see.

B. Customized production of industrial coatings should become one of our core activities in the next years. This exhibition sounds like a good opportunity to follow-up recent trends. I think it's best if we schedule our meeting after the exhibition. You'll input at our meeting is bound to be important.

C. I don't think much of these exhibitions. In my experience, you'll mostly find publicity stands at this exhibition. You'll probably won't learn anything you don't already know. So, I think it's better that our meeting takes place as originally scheduled.

D. I understand that this exhibition might be of particular interest to you and therefore, you'd like to be present. I'll reschedule the meeting to next week (26th). The production supervisors will recognize the importance of this exhibition and will agree to reschedule.

References

Ahmed, Y., Payne, T. and Whiddett, S. (1997) A Process for Assessment Exercise Design: A model of best practice. International Journal of Selection and Assessment, 5, 62-68.

Anseel, F. and Lievens, F. (2006) Certainty as a Moderator of Feedback Reactions? A Test of the Strength of the Self-Verification Motive. Journal of Occupational and Organizational Psychology, 79, 533-551.

Armstrong, R.D., Jones, D.H. and Wang, Z. (1994) Automated Test Construction Using Classical Test Theory. Journal of Educational Statistics, 19, 73-90.

Arthur, W., Jr, Day, E.A., McNelly, T.L. and Edens, P.S. (2003) A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions. Personnel Psychology, 56, 125-154.

Brannick, M.T., Michaels, C.E. and Baker, D.P. (1989) Construct Validity of In-Basket Scores. Journal of Applied Psychology, 74, 957-963.

Clause, C.C., Mullins, M.E., Nee, M.T., Pulakos, E.D. and Schmitt, N. (1998) Parallel Test Form Development: A procedure for alternative predictors and an example. Personnel Psychology, 51, 193-208.

Gibson, W.M. and Weiner, J.A. (1998) Generating Random Parallel Test Forms Using CTT in a Computer-Based Environment. Journal of Educational Measurement, 35, 297-310.

Hunter, J.E. and Schmidt, F.L. (2004) Methods of Meta-Analysis (2nd edn). Thousand Oaks, CA: Sage.

Irvine, S.H. and Kyllonen, P.C. (2002) Item Generation for Test Development. Mahwah, NJ: Lawrence Erlbaum Associates.

Lievens, F., Harris, M.M., Van Keer, E. and Bisqueret, C. (2003) Predicting Cross-Cultural Training Performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. Journal of Applied Psychology, 88, 476-489.

Lievens, F. and Sackett, P.R. (2007) Situational Judgment Tests in High Stakes Settings: Issues and strategies with generating alternate forms. Journal of Applied Psychology, 92, 1043-1055.

Nunnally, J.C. and Bernstein, I.H. (1994) Psychometric Theory (3rd edn). New York: McGraw Hill.

Schippmann, J.S., Prien, E.P. and Katz, J.A. (1990) Reliability and Validity of In-Basket Performance Measures. Personnel Psychology, 43, 837-859.

Stanton, J.M. and Rogelberg, S.G. (2001) Using Internet/Intranet Web Pages to Collect Organizational Research Data. Organizational Research Methods, 4, 200-217.

Tett, R.P., Guterman, H.A., Bleier, A. and Murphy, P.J. (2000) Development and Content Validation of a "Hyperdimensional" Taxonomy of Managerial Competence. Human Performance, 12, 232-251.

Tett, R.P. and Jackson, D.N. (1990) Organization and Personality Correlates of Participative Behaviours Using an in-Basket Exercise. Journal of Occupational Psychology, 63, 175-188.

Tett, R.P., Menard, D., Guterman, H. and Beauregard, R.S. (2001). Challenges in Computerizing Work Sample Measures: Testing the equivalence of an in-basket exercise. Paper Presented at the 16th Conference of the Society for Industrial and Organizational Psychology, San Diego.

Thornton, G.C. III and Mueller-Hanson, R. (2004) Developing Organizational Simulations: A guide for practitioners and students. Mahwah, NJ: Lawrence Erlbaum.

Woehr, D.J. and Arthur, W. Jr. (2003) The Construct-Related Validity of Assessment Center Ratings: A review and meta-analysis of the role of methodological factors. Journal of Management, 29, 231-258.